

TESTING STATIONARITY WITH SURROGATES — A ONE-CLASS SVM APPROACH

Jun Xiao, Pierre Borgnat, Patrick Flandrin

École Normale Supérieure de Lyon
46 allée d'Italie 69364 Lyon Cedex 07 France

Cédric Richard

Université de Technologie de Troyes
12 rue Marie Curie 10010 Troyes Cedex France

ABSTRACT

An operational framework is developed for testing stationarity relatively to an observation scale, in both stochastic and deterministic contexts. The proposed method is based on a comparison between global and local time-frequency features. The originality is to make use of a family of stationary surrogates for defining the null hypothesis and to base on them a statistical test implemented as a one-class Support Vector Machine. The time-frequency features extracted from the surrogates are considered as a learning set and used to detect departure from stationarity. The principle of the method is presented, and some results are shown on typical models of signals that can be thought of as stationary or nonstationary, depending on the observation scale used.

Index Terms— Stationarity Test, Time-Frequency Analysis, Support Vector Machines, One-Class Classification

1. REVISITING STATIONARITY

Considering *stationarity* is central in many signal processing applications, either because its assumption is a pre-requisite for applying most of standard algorithms devoted to steady-state regimes, or because its breakdown conveys specific information in evolutive contexts. Testing for stationarity is therefore an important issue, but addressing it raises some difficulties. The main reason is that the concept itself of “stationarity”, while uniquely defined in theory, is often interpreted in different ways. Indeed, whereas the standard definition of stationarity refers only to *stochastic processes* and concerns the invariance of statistical properties over time, stationarity is also usually invoked for *deterministic signals* whose spectral properties are time-invariant. Moreover, while the underlying invariances (be they stochastic or deterministic) are supposed to hold in theory for *all* times, common practice allows them to be restricted to some *finite* time interval [1, 2, 3], possibly with abrupt changes in between [4, 5]. As an example, we can think of speech that is routinely “segmented into stationary frames”, the “stationarity” of voiced segments relying in fact on periodicity structures within restricted time intervals. Those remarks call for a better framework aimed at dealing with “stationarity” in an operational sense, with a definition

that would both encompass stochastic and deterministic variants, and include the possibility of its test relatively to a given observation scale. This is the purpose of the present study.

2. FRAMEWORK

2.1. A time-frequency approach

As far as only second order evolutions are to be tested, time-frequency (TF) distributions and spectra are natural tools [6]. Well-established theories exist for justifying the choice of a given TF representation. In the case of stationary processes, the Wigner-Ville Spectrum (WVS) is not only constant as a function of time but also equal to the Power Spectrum Density (PSD) at each instant. From a practical point of view, the WVS is a quantity that has to be estimated. In this study, we choose to make use of multitaper spectrograms [7] defined as

$$S_{x,K}(t, f) = \frac{1}{K} \sum_{k=1}^K S_x^{(h_k)}(t, f),$$

where the $\{S_x^{(h_k)}(t, f), k = 1, \dots, K\}$ stand for the K spectrograms computed with the K first Hermite functions as short-time windows $h_k(t)$:

$$S_x^{(h_k)}(t, f) = \left| \int x(s) h_k(s - t) e^{-i2\pi f s} ds \right|^2.$$

The reason for this choice is that spectrograms can be both interpreted as estimates of the WVS for stochastic processes and as reduced interference distributions for deterministic signals. The multitaper approach is furthermore adopted in order to reduce estimation variance without some extra time-averaging which would be inappropriate in a nonstationary context. In practice, the multitaper spectrogram is evaluated only at N time positions $\{t_n, n = 1, \dots, N\}$, with a spacing $t_{n+1} - t_n$ which is an adjustable fraction of the temporal width T_h of the K windows $h_k(t)$.

2.2. Relative stationarity

The TF interpretation suggesting that suitable representations should undergo no evolution in stationary situations, stationarity tests can be envisioned on the basis of some comparison

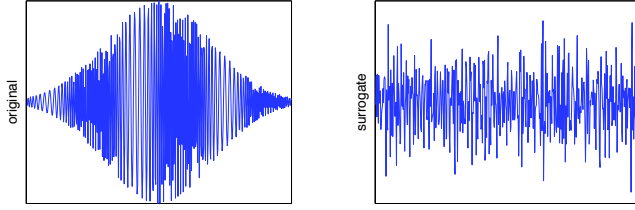


Fig. 1. Signal and surrogates. This figure displays a Frequency Modulated signal with a Gaussian amplitude, which is a nonstationary signal (left), and one of its surrogates obtained by replacement of the phase of the Fourier transform by an i.i.d. uniform phase (right).

between local and global features. Relaxing the assumption that stationarity would be some *absolute* property, the basic idea underlying the approach proposed here is that, when considered over a given duration, a process will be referred to as *stationary relatively to this observation scale* if its time-varying spectrum undergoes no evolution or, in other words, if the local spectra at all different time instants are statistically similar to the global spectrum obtained by marginalization.

3. TEST

3.1. Surrogates

Revisiting stationarity within the TF perspective has already been pushed forward [2], but the novelty is to address the significance of the difference “local vs. global” by elaborating from the data itself a *stationarized* reference serving as the null hypothesis for the test. Indeed, distinguishing between stationarity and nonstationarity would be made easier if, besides the analyzed signal itself, we had at our disposal some reference having the same marginal spectrum while being stationary. Since such a reference is generally not available, one possibility is to *create it from the data*: this is the rationale behind the idea of “surrogate data”, a technique which has been introduced and widely used in the physics literature, mostly for testing linearity [8, 9] (up to some proposal reported in [10], it seems to have never been used for testing stationarity).

For an identical marginal spectrum over the same observation interval, *nonstationary* processes are expected to differ from *stationary* ones by some structured organization in time, hence in their time-frequency representation. A set of J “surrogates” is thus computed from a given observed signal $x(t)$, so that each of them has the same PSD as the original signal while being “stationarized”. In practice, this is achieved by destroying the organized phase structure controlling the nonstationarity of $x(t)$, if any. To this end, $x(t)$ is first Fourier transformed to $X(f)$, and the modulus of $X(f)$ is then kept unchanged while its phase is replaced by a random one, uniformly distributed over $[-\pi, \pi]$. This modified spectrum is then inverse Fourier transformed, leading to as many station-

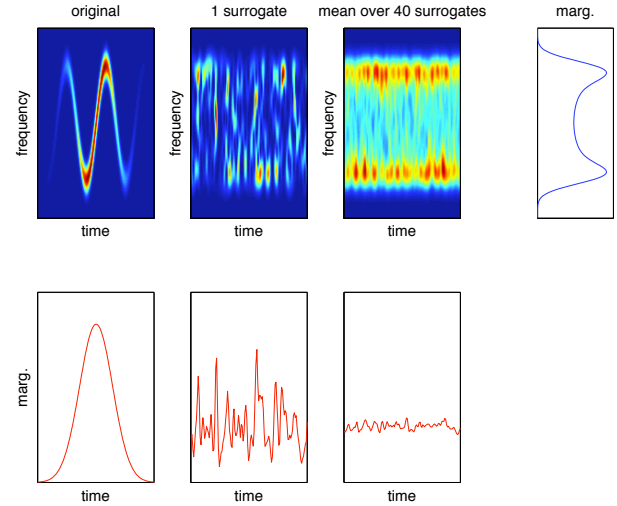


Fig. 2. Surrogates. This figure compares the TF structure of the nonstationary FM signal of Fig. 1 (1st column), of one of its surrogates (2nd column) and of the mean over $J = 40$ surrogates (3rd column). The spectrogram is represented in each case on the 1st line, with the corresponding marginal in time on the 2nd line. The marginal in frequency, which is the for the three spectrograms, is displayed on the far right of the 1st line.

ary surrogate signals as phase randomizations are operated. Fig. 1 shows a nonstationary signal and one surrogate resulting from this operation. The effect of the surrogate procedure is further illustrated in Fig. 2, displaying both signal and surrogate spectrograms, together with their marginals in time and frequency. It clearly appears from this figure that, while the original signal undergoes a structured evolution in time, the recourse to phase randomization in the Fourier domain ends up with stationarized (i.e., time unstructured) surrogate data with identical spectrum.

3.2. One-class SVM

Once a collection of stationarized surrogate data has been synthesized, different possibilities are offered. The first one is to extract from them some features such as *distances* between local and global spectra, and to characterize the null hypothesis of stationarity by the statistical distribution of their variation in time. This approach is the subject of current investigations that will be reported elsewhere [11]. We will here rather focus on an alternative viewpoint rooted in statistical learning theory: the collection of surrogates will be considered as a *learning set* and used to detect departure from stationarity. In this context, the classification task is fundamentally a one-class classification problem and differs from conventional two-class pattern recognition problems in the way how

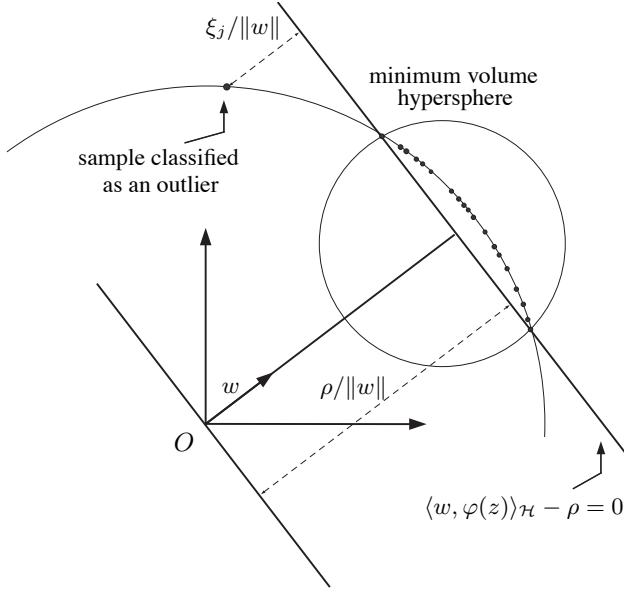


Fig. 3. One-class SVM with kernel $\kappa(z_i, z_j)$ depending only on $z_i - z_j$.

the classifier is trained. The latter uses only target data to estimate a boundary which encloses most of them. The machinery of one-class Support Vector Machines (1-class SVM), which was introduced for outlier detection [12], can be used. This technique has been successfully applied to a number of problems, including audio and biomedical signal segmentation [4, 13].

Let $\mathcal{Z} = \{z_1, \dots, z_J\}$ be a set of J surrogate signals (or a collection of features derived from it). Let $\kappa : \mathcal{Z} \times \mathcal{Z} \rightarrow \mathbb{R}$ be a kernel function that satisfies Mercer conditions. The latter can be used to map the z_j 's into a feature space denoted by \mathcal{H} via $\varphi : \mathcal{Z} \rightarrow \mathcal{H}$ defined as $\varphi(z) = \kappa(z, \cdot)$. The space \mathcal{H} is shown to be a reproducing kernel Hilbert space of functions with dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. The reproducing kernel property states that $\langle \kappa(z_i, \cdot), \kappa(z_j, \cdot) \rangle_{\mathcal{H}} = \kappa(z_i, z_j)$, which means that $\kappa(z_i, z_j)$ can be interpreted as the dot product between z_i and z_j mapped to \mathcal{H} by $\varphi(\cdot)$. A classic example of Mercer kernel is the Gaussian kernel defined as $\kappa(z_i, z_j) = \exp(-\|z_i - z_j\|^2 / 2\sigma_0^2)$, where σ_0^2 is a bandwidth parameter. Note that it maps any data point onto a hypersphere of radius 1 since $\kappa(z_j, z_j) = 1$ for all z_j .

The learning strategy adopted by 1-class SVM is to map the data into the feature space corresponding to the kernel function, and determine the hyperplane $\langle w, \varphi(z) \rangle_{\mathcal{H}} - \rho = 0$ which separates them from the origin with maximum margin. The decision function $d(z) = \text{sgn}(\langle w, \varphi(z) \rangle_{\mathcal{H}} - \rho)$ then gives on which side of the hyperplane any new point z falls in feature space, and determine if it may be considered as an outlier. For kernels $\kappa(z_i, z_j)$ depending only on $z_i - z_j$ such as the Gaussian kernel, which map data onto a hypersphere, this strategy is equivalent to finding the minimum volume hy-

persphere enclosing the data [14]; See Fig. 3. Now, let us focus on the optimization problem solved to get the hyperplane parameters w and ρ . On the one hand, the distance $\rho/\|w\|$ that separates the hyperplane from the origin must be maximized. But on the other hand, the number of target samples wrongly classified as outliers must be minimized. Such samples z_j satisfy inequalities of the form $\langle w, \varphi(z_j) \rangle_{\mathcal{H}} \geq \rho - \xi_j$ with $\xi_j > 0$. Based on these results, the decision function is found by minimizing the weighted sum of a regularization term $\|w\|^2$, and an empirical error term depending on the margin variable ρ and individual errors ξ_j

$$\begin{aligned} \min_{w, \rho, \xi} \quad & \frac{1}{2} \|w\|^2 + \frac{1}{\nu J} \sum_{j=1}^J \xi_j - \rho \\ \text{subject to} \quad & \langle w, \varphi(z_j) \rangle_{\mathcal{H}} \geq \rho - \xi_j, \quad \xi_j \geq 0, \end{aligned}$$

with $\nu \in [0, 1]$. Basic properties of 1-class SVM are reported in [12]. An important result is that the ν parameter may be used to incorporate prior information about the frequency of novelty occurrences.

We shall now use 1-class SVM with (stationary) surrogate signals. The resulting decision rules will allow us to distinguish between stationary and nonstationary processes.

4. TWO EXAMPLES

Two test signals are used as being simple illustrations of possible nonstationary evolutions: amplitude modulation (AM) of a random noise and deterministic frequency modulation (FM). The models are expressed, for $t \in [0, T]$, as

$$\begin{aligned} \text{(AM)} \quad & x(t) = (1 + \alpha \sin 2\pi t/T_0) e(t); \\ \text{(FM)} \quad & x(t) = \sin 2\pi(f_0 t + \alpha \sin 2\pi t/T_0) + e(t), \end{aligned}$$

where $e(t)$ is white Gaussian noise and (FM case) f_0 is the central frequency. For both models, T_0 is the period of the modulation and $0 \leq \alpha \leq 1$ is the modulation factor. All non-stationarities cannot be subsumed under these two categories, but they are believed to give meaningful examples.

Given $x(t)$, TF features are extracted from a multitaper spectrogram. Introducing first the local average $\langle \cdot \rangle_{\tilde{s}_n}$ computed with the normalized time-frequency distribution (for $f > 0$ only):

$$\tilde{S}_n(f) := \frac{S_{x,K}(t_n, f)}{\int_0^\infty S_{x,K}(t_n, f) df}$$

at times t_n ($n = 1, \dots, N$), temporal features are obtained so as to describe the time evolution of the local power P_n of the signal and of its local frequency content F_n :

$$P_n = \langle 1 \rangle_{\tilde{s}_n}; \quad F_n = \langle f \rangle_{\tilde{s}_n}; \quad F_n^2 = \langle f^2 \rangle_{\tilde{s}_n}$$

From this (keeping a small number of features for a sake of clarity), we retain the following two characteristics comparing local TF features to global ones:

$$\begin{cases} P = \text{std}(\{P_n\}_{n=1..N}) / \text{mean}(\{P_n\}_n) \\ F = \text{std}(\{F_n\}_{n=1..N}) / \text{mean}(\{\sqrt{F_n^2 - (F_n)^2}\}_n) \end{cases}$$

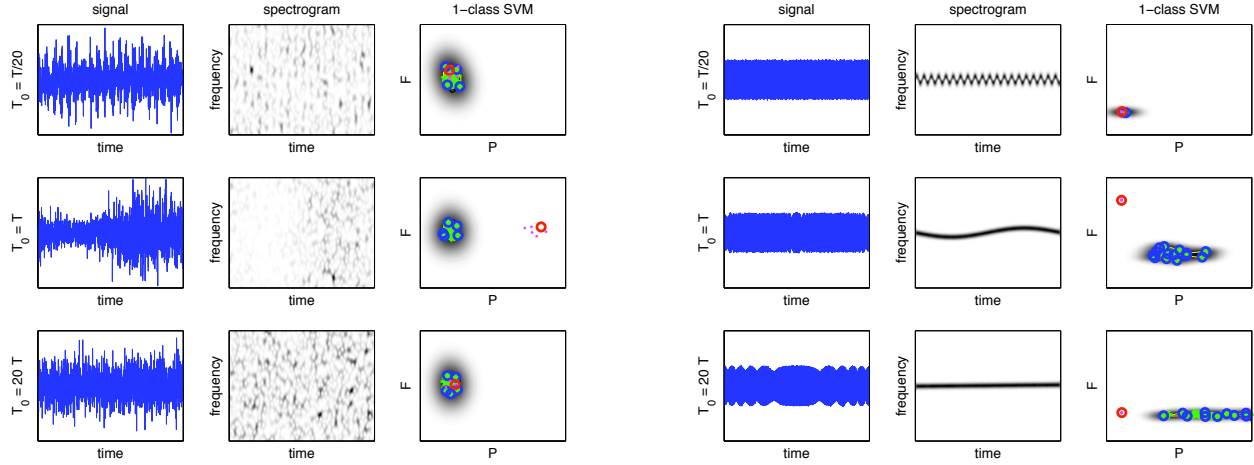


Fig. 4. Two examples. Signals, spectrograms and space (P, F) of the TF features in AM (left) and FM (right) situations. From top to bottom, $T_0 = T/20, T$ and $20T$, with $T = 1600$. In each case, the red circle corresponds to the (P, F) pair of one test signal used to draw the surrogates. Those surrogates ($J = 40$ in the experiments reported here) are plotted as green dots which, with 1-class SVM, define the domain of stationarity represented here as the gray shaded region, the blue circles corresponding to the support vectors. Magenta dots are independent realizations of the same test model. Other parameters are as follows — number of tapers: $K = 5$, length of tapers: $T_h = 387$, modulation indices: $\alpha = 0.5$ (AM) and 0.02 (FM), signal-to-noise ratio: SNR = 10 dB — SVM kernel κ is Gaussian, $\sigma = 0.07, \nu = 0.05$.

The first one (P) is a measure of the fluctuations in time of the local power of the signal, whereas the second one (F) operates the same way with respect to the local mean frequency. These characteristics are used as features, $z = (P, F)$, for the 1-class SVM whose output is displayed in Fig. 4. The SVM toolbox proposed in [15] was used for this illustration.

The results are shown for $T_0 = T/20, T$ and $20T$, allowing to consider stationarity relatively to the ratio between the observation time T and the modulation period T_0 . They can be summarized as follows;

Macroscale — For a small modulation period (or a large observation time, i.e., when $T_0 \ll T$), the situation can be considered as stationary, due to the observation of many similar oscillations over the observed time scale. This is reflected by a test signal (P, F) feature (red circle, see caption) which lies inside the region defined by the 1-class SVM for the stationary surrogates.

Mesoscale — For a medium observation time ($T \approx T_0$), the local evolution due to the modulation is prominent and the red circle for the modulated signal is well outside the stationary region, in accordance with a situation that can be referred to as nonstationary.

Microscale — Finally, if $T_0 \gg T$, the result turns back to stationarity because no significative change in the amplitude or the frequency is observed over the considered time scale.

Two remarks can be made with respect to these results. First, although both AM and FM signals are seen as nonstationary in the mesoscale regime, in the AM case the nonstationarity manifests through a deviation of the local power P , whereas in the FM case, it is the local frequency F that is mostly different from the stationary class. Second, it turns out that, in the microscale regime, the deterministic stationarity (FM case) naturally ends up with a much larger dispersion in P than in F since, by construction, the spectrum is narrowband. Moreover, the randomization which underlies the construction of surrogates necessarily ends up with more power fluctuations in the stationarized data than in the original test signal, and hence with a (P, F) pair which, at best, lies on the border of the support of the stationary class. This suggests that, in some sense, the position of the (P, F) pair with respect to the stationary region does not only give an information about a possible nonstationarity but also an indication about its type.

5. CONCLUSION

Testing for stationarity in signal processing and data analysis has already received some attention, but maybe not as much as might be expected from its ubiquitous nature. In this paper, we proposed an operational framework to measure and test departures from stationarity. Its originality is that it makes use of a family of stationarized realizations of the analyzed signal, called surrogates, for defining the null

hypothesis. Time-frequency features are then extracted from surrogates, and used as a learning set to train a 1-class SVM which encompasses what may be considered stationary.

A number of extensions to the present work are possible. Here, we use in Section 4 specific bidimensional features, but one can think of directly using general TF representations emerging from the use of SVM machinery for Time-Frequency, such as in [16]. Another possibility is to generalize the present approach to other forms of stationarity. This requires to define new specific stationarizing tools and signal representations, in the spirit of [17, 18].

6. REFERENCES

- [1] S. Mallat, G. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. of Stat.*, vol. 24, no. 1, pp. 1–47, 1998.
- [2] W. Martin and P. Flandrin, "Detection of changes of signal structure by using the Wigner-Ville spectrum," *Signal Proc.*, vol. 8, pp. 215–233, 1985.
- [3] R.A. Silverman, "Locally stationary random processes," *IRE Trans. on Info. Theory*, vol. 3, pp. 182–187, 1957.
- [4] M. Davy and S. Godsill, "Detection of abrupt signal changes using Support Vector Machines: An application to audio signal segmentation," in *Proc. IEEE ICASSP-02*, Orlando (FL), 2002.
- [5] H. Laurent and C. Doncarli, "Stationarity index for abrupt changes detection in the time-frequency plane," *IEEE Signal Proc. Lett.*, vol. 5, no. 2, pp. 43–45, 1998.
- [6] P. Flandrin, *Time-Frequency / Time-Scale Analysis*, Academic Press, 1999.
- [7] M. Bayram and R.G. Baraniuk, "Multiple window time-varying spectrum estimation," in *Nonlinear and Non-stationary Signal Processing*, W.J. Fitzgerald et al., Ed. 2000, Cambridge Univ. Press.
- [8] J. Theiler et al., "Testing for nonlinearity in time series: the method of surrogate data," *Physica D*, vol. 58, no. 1–4, pp. 77–94, 1992.
- [9] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Phys. Rev. Lett.*, vol. 77, no. 4, pp. 635–638, 1996.
- [10] C.J. Keylock, "Constrained surrogate time series with preservation of the mean and variance structure," *Phys. Rev. E*, vol. 73, pp. 030767.1–030767.4, 2006.
- [11] J. Xiao, P. Borgnat, and P. Flandrin, "Testing stationarity with time-frequency surrogates," in *Proc. EUSIPCO-07*, Poznan (PL), Sept. 2007, to appear.
- [12] B. Schölkopf, J.C. Platt, J.S. Shawe-Taylor, A.J. Smola, and R.C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [13] A.B. Gardner, A.M. Krieger, G. Vachtsevanos, and B. Litt, "One-class novelty detection for seizure analysis from intracranial eeg," *Journal of Machine Learning Research*, vol. 7, pp. 1025–1044, 2006.
- [14] D.M.J. Tax and R.P.W. Duin, "Support vector data description," *Machine Learning*, vol. 54, pp. 45–66, 2004.
- [15] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and Kernel Methods Matlab Toolbox," Perception Systèmes et Information, INSA de Rouen, <http://asi.insa-rouen.fr/~arakotom/toolbox/index.html>, 2005.
- [16] P. Honeine, C. Richard, and P. Flandrin, "Time-frequency learning machines," *IEEE Trans. on Signal Proc.*, vol. 55, no. 7 (Part 2), pp. 3930–3936, 2007.
- [17] P. Flandrin, P. Borgnat, and P.-O. Amblard, "From stationarity to self-similarity, and back : Variations on the Lamperti transformation," in *Processes with Long-Range Correlations: Theory and Applications*, G. Ranganjarian and M. Ding, Eds. June 2003, vol. 621 of *Lectures Notes in Physics*, pp. 88–117, Springer-Verlag.
- [18] P. Borgnat, P.-O. Amblard, and P. Flandrin, "Stochastic invariances and Lamperti transformations for stochastic processes," *J. Phys. A: Math. Gen.*, vol. 38, no. 10, pp. 2081–2101, Feb. 2005.